

This is the authors' accepted manuscript for the article:

Pinar, Mehmet, Milla, Joniada, and Stengos, Thanasis (2018) Sensitivity of university rankings: Implications of stochastic dominance efficiency analysis. Education Economics, forthcoming, <https://doi.org/10.1080/09645292.2018.1512560>

The final publication is available via:

<https://www.tandfonline.com/doi/full/10.1080/09645292.2018.1512560>

# Sensitivity of university rankings: Implications of stochastic dominance efficiency analysis\*

Mehmet Pinar<sup>†</sup>

Joniada Milla<sup>‡</sup>

Thanasis Stengos<sup>§</sup>

August 3, 2018

## Abstract

To create their rankings, university-ranking agencies usually combine multiple performance measures into a composite index. However, both rankings and index scores are sensitive to the weights assigned to performance measures. This paper uses a stochastic dominance efficiency methodology to obtain two extreme, case-weighting vectors using the Academic Ranking of Worldwide Universities (ARWU) and Times Higher Education (THE) data, both of which lead to the highest and lowest index outcomes for the majority of universities. We find that both composite scores and rankings are very sensitive to weight variations, especially for middle- and low-ranked universities.

*JEL Classifications:* C12; C13; C14; C15; I23

*Key Words:* University rankings; Higher education; Nonparametric Stochastic Dominance Efficiency; Mixed Integer Programming

---

\*We thank Brian Hotson for all his efforts in editing our manuscript.

<sup>†</sup>Corresponding author: Business School, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, United Kingdom; e-mail: mehmet.pinar@edgehill.ac.uk

<sup>‡</sup>Department of Economics, Sobey School of Business, Saint Mary's University, Halifax, NS B3H 3C3, Canada; e-mail: joniada.milla@smu.ca

<sup>§</sup>Department of Economics, University of Guelph, N1G 2W1, Guelph, Ontario, Canada; e-mail: tsten-gos@uoguelph.ca

# 1 Introduction

University rankings have attracted policy makers’ attention, and have led to major changes in national policies such as encouraging mergers of higher education institutions (HEIs) (Docampo et al., 2015) and promoting differentiation of research-intensive HEIs from teaching-intensive universities.<sup>1</sup> Rankings also affect student university choice (Griffith and Rask, 2007; Horstschräer, 2012; Gibbons et al., 2015; Broecke, 2015) given the significant differences in earnings between top university graduates compared to those of lower ranked institutions (see Milla, 2017 for a recent review). The impact of rankings at a public-policy, institution<sup>2</sup>, and individual level have triggered an increasing debate about the way in which the rankings are constructed. We contribute to this debate by testing the sensitivity of university rankings.

University rankings are generally obtained by combining a university’s characteristics into a composite index. These characteristics — or performance criteria — are assigned weights by ranking providers to reflect their relative importance in assessing a university’s overall performance. However, there is an important caveat related to the assignment of weights; ranking authorities use their discretion to determine how to weight each university-performance criteria based on their own value judgments (Kehm, 2014). As such, even a relatively minor change in weight allocation may lead to substantial alterations in rankings (see e.g., Saisana et al., 2011). This may lead to serious implications in both university- and national-level policy-related projections when the resulting rankings are used for high-stakes decision-making. Assessing the way these rankings are constructed and their sensitivity with respect to weight allocation, then, becomes key. A better ranking index is not the product of this paper, but evaluation of the full grid of weight vectors and a provision of insights on the interpretation and differences that these rankings embed.

Finding an optimal weighting scheme that ensures a comprehensive and fair evaluation for all institutions is a challenging task because the true objective function underlying a “correct” ranking of university quality is unknown. Even though alternative-weighting allocations across

---

<sup>1</sup>Hazelkorn (2011, 2014) provides a detailed discussion about policy and institutional changes after the world university rankings.

<sup>2</sup>For instance, Cyrenne and Grant (2009) find that different types of Canadian universities follow different methods to raise their reputation ranking.

performance criteria are employed in previous literature<sup>3</sup>, testing all possible weight allocation across performance criteria has been neglected in the literature due to the computational difficulties and the voluminous nature of this analysis. We employ the *stochastic dominance efficiency* (SDE) approach in a novel way to conduct this analysis. The method was first adopted by Scaillet and Topaloglou (2010) as a procedure to find a weight vector that produces the most efficient market portfolio of financial assets by maximizing portfolio returns. SDE is a direct extension of the pair-wise stochastic dominance (SD) analysis to a multivariate framework.<sup>4</sup> Even though pair-wise SD tests were used by the literature to compare quality of universities based on single performance measures<sup>5</sup>, SDE tests have not been employed to test the sensitivity of composite scores and rankings based on alternative-weight allocations across performance criteria, which this paper aims to do.

In the context of our paper, we may analogously think of quality of universities to be characterized by a multivariate portfolio of assets, which can be university characteristics or performance criteria. The SDE approach selects a set of weights that produces an index whose distribution stochastically dominates the distributions of all other possible indices. Applied to university data, SDE delivers an index of university performance that stochastically dominates all other possible indices constructed with the same set of university performance criteria. As a result, the performance measures in which the majority of universities perform better are weighted relatively more in this index.

The percentage of universities that make up the “majority” is endogenously determined by the available data and the SDE procedure. Hereafter we refer to this set of weights as the *Highest*

---

<sup>3</sup>For instance, Saisana et al. (2011) use three alternative weighting (i.e., factor analysis derived weights, equal weighting and “university-specific weighting” that maximizes that university’s performance relative to all other universities).

<sup>4</sup>The most popular use of the pair-wise SD analysis in the literature is to determine which population is better in terms of a given well-being dimension compared to the other (see e.g., Atkinson, 1970; Shorrocks, 1983; Kakwani, 1984; Atkinson, 1987; Foster and Shorrocks, 1988; Ravallion, 1994; Davidson and Duclos, 2000; Barrett and Donald, 2003; Agliardi et al., 2017 among many others). This SD comparisons has moved to a multivariate one by analyzing various welfare dimensions and portfolios (see e.g., Post, 2003; Kuosmanen, 2004; Linton et al., 2005; Duclos et al., 2006; Agliardi et al., 2012; Pinar et al., 2013; Delgado and Escanciano, 2013; Agliardi et al., 2014; Gonzalo and Olmo, 2014; Linton et al., 2014; Yalonetzky, 2014; Agliardi et al., 2015; Pinar, 2015; Pinar et al., 2015; Pinar et al., 2017 among many others).

<sup>5</sup>For instance, Bazen and Moyes (2012), and Carayol and Lahatte (2014) use pair-wise SD tests to compare the distribution of publication performance of staff members to rank these institutes. In their application, both the quality and quantity of publications are taken into account and all possible pairs of institutes are compared based on the publication performance.

*Index for Most* (HIM) weight vector. HIM is one polar-weight allocation scenario that we consider where empirical distribution of composite scores obtained by universities dominates other empirical distributions of composite scores obtained with any other alternative-weight allocation across performance criteria. The other we call the *Least Index for Most* (LIM) weight vector. The LIM vector is retrieved by adjusting the SDE procedure to deliver the index whose distribution is stochastically dominated by all other possible indices. In this index, the performance measures for which the majority of universities perform poorly receive relatively more weight. Once we attain the HIM and LIM weight vectors, we make a comparison. This empirical exercise tests the sensitivity of the rankings when performance criteria are weighted differently, and which universities are mostly affected. As a result, governments, institutions and individuals that intend to incorporate rankings in policy design, budget reallocation and decision-making, will have clearer empirical means to evaluate how sensitive the expected results can be depending on the rankings' weight vector.

The data we use for this analysis are publicly available on the ARWU (Academic Ranking of World Universities) and the THE (Times Higher Education) world university rankings webpages.<sup>6</sup> We find that the variables weighted heavily in HIM and LIM remain consistently the same over time for both university and department rankings. Based on three scenarios (benchmark, HIM and LIM), we find that composite scores and rankings of universities in the ARWU and THE are extremely sensitive to alternative-weight allocations across performance criteria. Only a handful of universities experience low levels of variation in their composite scores and rankings. These universities are ranked in the top 10 in ARWU and top 40 in THE rankings. Furthermore, we also find a strong negative correlation between composite score volatility (i.e., coefficients of variation of benchmark, HIM, and LIM scenarios) and the benchmark rankings suggesting that the universities ranked at the top (middle and low) in benchmark ranking experience a relatively lower (higher) variation in their composite scores. Overall, our findings raise awareness that composite scores and rankings of universities are significantly sensitive to differential weight allocations across performance criteria; they should be approached with caution.

---

<sup>6</sup>Please refer to <https://www.timeshighereducation.co.uk/world-university-rankings/> for the THE World University Ranking data and <http://www.shanghairanking.com/> for ARWU rankings data (also known as Shanghai ranking).

The organization of the paper is as follows: Section 2 describes the SDE methodology; Section 3 describes the data; and Section 4 presents the results. Section 5 concludes.

## 2 Methodology

### 2.1 SDE Approach

In this section we summarize the SDE methodology. We use  $\mathbf{Y}$  to represent an  $m \times N$  normalized achievement matrix of  $m$  performance criteria for  $N$  universities that is defined in  $\mathbb{R}^m$ . We denote by  $F(\mathbf{y})$ , the continuous cumulative distribution function (cdf) of  $\mathbf{Y} = (Y_1, \dots, Y_m)'$  at point  $\mathbf{y} = (y_1, \dots, y_m)'$ . Using a vector of predetermined weights,  $\mathbf{w}$ , ranking agencies construct a composite index for all of the universities (i.e.,  $\mathbf{wY}$ ). Then these composite indices are used by the rankers to produce university rankings that we refer to as benchmark ranking.

Let us consider an alternative weighting vector  $\mathbf{w}_a \in \mathbb{L}$  where  $\mathbb{L} := \{\mathbf{w}_a \in \mathbb{R}_+^m : \mathbf{e}'\mathbf{w}_a = 1\}$  and  $\mathbf{e}$  is a vector of ones. The condition  $\mathbf{e}'\mathbf{w}_a = 1$  ensures that all indicators are assigned non-negative weights and that these weights sum to one. An ‘alternative’ index (and related ranking) can be built by using  $\mathbf{w}_a$  and  $\mathbf{Y}$ . Let us further denote by  $G(s, \mathbf{w}; F)$  and  $G(s, \mathbf{w}_a; F)$  the cdfs of the benchmark and alternative indices (i.e.,  $\mathbf{w}'\mathbf{Y}$  and  $\mathbf{w}_a'\mathbf{Y}$  respectively) at point  $s$  given by  $G(s, \mathbf{w}; F) := \int_{\mathbb{R}^m} \mathbb{I}\{\mathbf{w}'\mathbf{u} \leq s\} dF(\mathbf{u})$  and  $G(s, \mathbf{w}_a; F) := \int_{\mathbb{R}^m} \mathbb{I}\{\mathbf{w}_a'\mathbf{u} \leq s\} dF(\mathbf{u})$ , respectively, where  $s$  represents an index score,  $\mathbb{I}$  is an indicator function, and  $\mathbf{u}$  is any increasing monotonic function of  $s$  such that  $\mathbf{u}'(s) > 0$  (Scaillet and Topaloglou, 2010). We apply the first-order SDE test to monotonically increasing functions of anonymous and additively decomposable social welfare functions of the form  $g(\mathbf{W}) = \int \mathbf{u}(s) d\mathbf{W}(s)$ , where  $g$  is a strictly increasing function, and  $\mathbf{W}$  is the distribution of weighted university-specific utility functions.

The general hypotheses for testing the first-order SDE of  $\mathbf{w}$  can be written as:

$$H_0 : G(s, \mathbf{w}; F) \leq G(s, \mathbf{w}_a; F) \quad \text{for all } s \in \mathbb{R} \text{ and for all } \mathbf{w}_a \in \mathbb{L},$$

$$H_1 : G(s, \mathbf{w}; F) > G(s, \mathbf{w}_a; F) \quad \text{for some } s \in \mathbb{R} \text{ or for some } \mathbf{w}_a \in \mathbb{L}.$$

Under the null hypothesis ( $H_0$ ), the distribution of the benchmark index is not bigger in magnitude than the distribution of alternative index. Therefore,  $G(s, \mathbf{w}; F)$  is smaller than  $G(s, \mathbf{w}_a; F)$  for all index scores,  $s$ . If the null hypothesis is not rejected by using any alternative-weighting vector to obtain index scores, this suggests that the benchmark index produces a distribution of composite scores where there is always a higher proportion of institutes that have a score that is above any given  $s$ .

Under the alternative hypothesis ( $H_1$ ), for some index score,  $s$ , some universities' scores in the alternative index are higher in magnitude than the ones in the benchmark index. Therefore,  $G(s, \mathbf{w}; F)$  is greater than  $G(s, \mathbf{w}_a; F)$  for some index score  $s$ , and the benchmark index is stochastically dominated by an alternative index at some  $s$ . This means that the alternative index provides a higher index score for some universities. The inequality in the definition of  $H_1$  above suggests that the proportion of universities with index scores less than  $s$  is smaller for the alternative index than for the benchmark index. This means that a higher proportion of universities have scores that are higher than  $s$  according to the alternative index (constructed using  $\mathbf{w}_a$ ) when compared to the benchmark index (constructed using  $\mathbf{w}$ ).

We obtain the empirical counterpart of  $F$  (i.e.  $\hat{F}$ ) by simply integrating the empirical distribution:

$$G(s, \mathbf{w}; \hat{F}) = \frac{1}{N} \sum_{n=1}^N \mathbb{I}\{\mathbf{w}'\mathbf{Y} \leq s\}$$

The empirical counterpart of  $\mathbf{w}_a$  can be obtained in a similar manner. The asymptotic distribution of  $\hat{F}$  is given by  $\sqrt{N}(\hat{F} - F)$  which tends weakly to a mean-zero Gaussian process  $\mathcal{B} \circ F$  in the space of continuous functions on  $\mathbb{R}^m$  (see e.g., the multivariate functional central limit theorem for stationary strongly mixing sequences stated in Rio (2000)).<sup>7</sup>

We use a weighted Kolmogorov-Smirnov type test statistic

$$\hat{S} := \sqrt{N} \frac{1}{N} \sup_{s, \mathbf{w}_a} \left[ G(s, \mathbf{w}; \hat{F}) - G(s, \mathbf{w}_a; \hat{F}) \right],$$

with decision rule:

---

<sup>7</sup>Scaillet and Topaloglou (2010) derive the limiting behavior by using the Continuous Mapping Theorem in their Lemma 2.1, similar to that of Lemma 1 of Barrett and Donald (2003).

“reject  $H_0$  if  $\hat{S} > c$ ”,

where  $c$  is the critical value. In order to make the result operational, we need to find an appropriate critical value,  $c$ . Since the distribution of the test statistic depends on the underlying distribution, this is not an easy task, and we decide to rely on a subsampling bootstrap method to stimulate p-values as proposed by Linton et al. (2005).

## 2.2 Mathematical formulation of the test statistics

The test statistic for the first-order SDE test is derived using a mixed-integer programming formulation that can be summarized as follows:

$$\max_{\mathbf{s}, \mathbf{w}_a} \hat{S} = \sqrt{N} \frac{1}{N} \sum_{n=1}^N (L_n - W_n) \quad (1a)$$

$$\text{s.t. } M(L_n - 1) \leq s - \mathbf{w}'\mathbf{Y} \leq ML_n, \quad \forall n \quad (1b)$$

$$M(W_n - 1) \leq s - \mathbf{w}'_a\mathbf{Y} \leq MW_n, \quad \forall n \quad (1c)$$

$$\mathbf{e}'\mathbf{w}_a = 1, \quad (1d)$$

$$\mathbf{w}_a \geq 0, \quad (1e)$$

$$W_n \in \{0, 1\}, L_n \in \{0, 1\}, \quad \forall n \quad (1f)$$

with  $M$  being a large constant. Mixed integer programming is used to maximize the distance between the sum of two binary variables,  $\frac{1}{N} \sum_{n=1}^N L_n$  and  $\frac{1}{N} \sum_{n=1}^N W_n$  which represent  $G(s, \mathbf{w}; \hat{F})$  and  $G(s, \mathbf{w}_a; \hat{F})$ , respectively, for every alternative weight vector  $\mathbf{w}_a$ . According to inequality (1b),  $L_n$  equals 1 for each scenario  $n \in N$  for which  $s \geq \mathbf{w}'\mathbf{Y}$ , and 0 otherwise. Similarly, inequality (1c) ensure that  $W_n$  equals 1 for each scenario for which  $s \geq \mathbf{w}'_a\mathbf{Y}$ , and 0 otherwise. Equation (1d) states that the sum of weights attached to performance criteria is one, while inequality (1e) ensures that the weights are non-negative. This formulation allows us to test the stochastic dominance of the benchmark index  $(\mathbf{w}'\mathbf{Y})$  over the alternative one  $(\mathbf{w}'_a\mathbf{Y})$ .

In the above maximization problem, binary variables are restricted by the two constraints



that define a lower- and an upper-bound. The solution to the problem is reached if bounds match. Otherwise, feasible regions are divided into sub-regions and sub-problems are solved. The algorithm is then applied to these sub-problems to obtain a feasible but not a global optimal solution. This procedure continues recursively for each sub-problem until all are solved or pruned. This procedure could take extensive amount of time to solve. However, the problem can be reformulated into a more tractable form to reduce the solving time. Given that we already know the distribution of the benchmark index (i.e.,  $\mathbf{w}'\mathbf{Y}$ ), rather than using all possible scores of  $s$  to obtain the alternative weight vector ( $\mathbf{w}_a$ ) that maximizes the distance, we can use the benchmark index for  $s$  scores. Once we fix  $s$  to  $\mathbf{w}'\mathbf{Y}$ , the  $\sum_{n=1}^N L_n$  can be easily calculated, which is the number of institutions with index scores equal or lower than  $s$ . As a result, we can reduce the optimization problem to only find the alternative weight vector that minimizes the number of institutions with an index score equal or lower than the  $s$  value ( $s \geq \mathbf{w}'_a\mathbf{Y}$ ). This would simply maximize the distance between the two distributions, and the solution could be obtained in a relatively shorter period of time (see section 4.1 of Scaillet and Topaloglou (2010) for the derivation of this formulation for the time series data).

In this section, we have described the methodology we use to produce the weighting vector for HIM scenario. The weighting vector for the LIM scenario can be obtained similarly by reversing the order of indices' cdfs in the model. We omit the details in favor of space.

### 3 Data and Descriptive Statistics

Data used in our analysis are obtained from ARWU and THE web pages. We examine the world-wide university and EB department rankings of ARWU, which are based on six and five performance criteria, respectively. The variables available in the ARWU data include: 1) number of the alumni of an institution winning Nobel Prizes and Fields Medals (Alumni hereafter); 2) number of the staff of an institution winning Nobel Prizes in Physics, Chemistry, Medicine and Economics and Fields Medal in Mathematics (Award hereafter); 3) number of staff members that are considered in the category of highly cited researchers in 21 broad subject categories (HiCi hereafter); 4) number of papers published in Nature and Science (N&S hereafter); 5) number of papers indexed

in Science Citation Index-Expanded and Social Science Citation Index (PUB hereafter); and 6) per capita academic performance of an institution (PCP hereafter), which is measured by the weighted scores of the other five indicators divided by the number of full-time equivalent academic staff.

Five variables are used to obtain the ARWU EB department rankings: 1) number of alumni of an institution winning Nobel prizes in Economics since 1961 (Alumni hereafter); 2) number of staff members that won a Nobel Prize in Economics since 1971 (Award hereafter); 3) number of staff members that are listed in the EB Category highly cited researchers (HiCi hereafter); 4) number of papers indexed in Social Science Citation Index in EB fields (PUB hereafter); and 5) percentage of papers published in top 20% journals of EB fields (TOP hereafter). Tables 1 and 2 summarize details about the performance criteria used for worldwide university and EB department rankings, and their respective weights.<sup>8</sup>

Our analysis for the ARWU university and EB department rankings covers the period between 2009 and 2015, and is conducted separately for each year because of the normalization of performance indicators.<sup>9</sup> The normalization of performance indicators used in ARWU is done by assigning a score of 100 to the highest scoring institution in a particular performance indicator, while another institutions' score is calculated as a percentage of the top-scoring institution for each performance indicator in a given year. Tables 3 and 4 show descriptive statistics for normalized performance indicators between 2009 and 2015. All (except for TOP in Table 4) are positively skewed but at a different level, which introduces varying score inequality for each indicator compared to the top-ranked institution.

We also provide another set of results using the THE university rankings data because the THE data variables are different from those in ARWU data. Table 5 provides performance-criteria details of THE world university rankings, and the associated pre-determined weight allocations. Compared to the ARWU rankings, not only do THE rankings use more indicators in their ranking, but they also include university performance aspects beyond research. THE world ranking data include five main performance criteria: 1) Teaching, 2) Research, 3) Citations, 4) Industry, and 5)

---

<sup>8</sup>Please refer to the <http://www.shanghairanking.com/> for the publicly available data set and detailed definition of each performance criteria used for the rankings.

<sup>9</sup>See Billaut et al., 2010 for a discussion on how normalization procedure of ARWU affects rankings, and that the overall scores cannot be compared over time.

International Outlook.

*Teaching* as a performance criteria contributes to the THE overall ranking with a 30% weight, and is composed of five indicators: perceived prestige of institutions in teaching from Thomson Reuters' reputation survey of worldwide poll of experienced scholars, staff-to-student ratio, ratio of doctoral to bachelor's degrees awarded by each institution, number of doctorates awarded by an institution (scaled against its size as measured by the number of academic staff it employs), measure of institutional income scaled against academic staff size.

*Research* as a performance criteria contributes 30% to the THE overall ranking and is a weighed mean of three indicators: a university's perceived reputation for research excellence among its peers (based on 10,000-plus responses to an annual academic reputation survey), university research income (scaled against staff numbers, and normalized for purchasing-power parity), and research output scaled against staff size.

*Citations* as a performance criteria measures the research influence of institutions by the number of citations of an institution's published research over time. This receives a weight of 30% towards the overall THE rankings.

*Industry* as a performance criteria measures the research income an institution earns from industry, which is a proxy for an institution's quality and impact on society. This variable contributes to the THE ranking with a 2.5% weight.

Finally, *International Outlook* as a performance criteria contributes 7.5% to the THE rankings. It is a weighted average of a university's ratio of international to domestic students, ratio of international to domestic staff, and proportion of a university's total research journal publications that have at least one international co-author.<sup>10</sup>

The THE data, differently from the ARWU data, contain information not only on research performance but also on financial and teaching performance of universities. Similar to ARWU data, the variables are measured in different metrics and then normalized.<sup>11</sup> The THE rankings

---

<sup>10</sup>It should be noted that in order to use all available variables in our analysis, we left out the institutions that lack data for some of performance indicators. In both rankings, ranking providers reallocate weights across other variables if some institutions have missing information for some indicators. However, in order to keep the pre-determined weights the same, we dropped institutions that have missing information for some criteria.

<sup>11</sup>For a detailed description of the THE world university rankings methodology, please refer to

are retrieved from the weighted average of the normalized variables using the pre-determined weights shown in Table 5.

We use the publicly available THE data between 2012 and 2015 on five performance criteria. Table 6 shows the descriptive statistics for these variables. On average, *Citations* is the highest in value, followed by *International Outlook*, *Industry*, *Teaching* and *Research*. The distributions of *Teaching*, *Research* and *Industry* are positively skewed. The distributions of *Citations* and *International Outlook* are almost normally distributed.

## 4 Results

### 4.1 ARWU University Rankings

We have access to two types of ARWU rankings and the associated data: the top-500 worldwide university rankings and the EB departmental rankings. We use the SDE methodology to find an alternative index for which most universities score their highest possible value — the HIM (highest index for most), and the weight vector associated with it, the HIM weights. Similarly, we use the SDE methodology to find an alternative index for which most universities score their lowest value — the LIM (lowest index for most) and the weight vector associated with it, the LIM weights. Our aim is to compare the relative weights between the benchmark weight vector, the HIM and LIM weight vectors, and examine the sensitivity of index composite scores and rankings. The objective is to find the source of sensitivity, based on the extent each performance criteria contributes in each ranking index (i.e. relative weights). This comparison can also be made over time to check whether the same performance criteria contribute more or less through years in a consistent way.

Table 7 displays the weight vectors for each year (2009–2015) for the HIM and LIM scenarios, respectively. Weight vectors in both cases show a similar trend over time. In the HIM scenario (Panel A), the number of publications (PUB) receives almost a full weight with a slight contribution from PCP. The other indicators receive a weight of zero (i.e., these indicators are left out when calculating the HIM index). This result suggests that the majority of the institutions are relatively

---

<https://www.timeshighereducation.com/world-university-rankings-2014-15-methodology>.

closer to the top ranked institution in terms of PUB and PCP when compared to other performance indicators. In the LIM scenario (Panel B), two variables that measure the number of Nobel prizes and Field Medals for alumni and faculty receive non-zero weights. The LIM weight vector provides an index whose distribution is stochastically dominated by any other index that can be constructed with the same variables but different set of weights. Both HiCi and N&S have a zero weight in both HIM and LIM scenarios because scores in these variables provide intermediate achievement levels for most institutions when compared to the other variables. Our findings suggest that if LIM variables (i.e., Alumni and Awards variables that get non-zero weights in the LIM scenario) are weighted more heavily, most of the universities fall behind the highest ranked institution. Similarly, if the HIM variables are weighted relatively more, most of the universities' index scores are relatively closer to the top ranked institution.

In order to easily see the differences between the benchmark, HIM and the LIM index, we plot in Figure 1 the cumulative distributions of the three indices for 2015. Index scores appear in the horizontal axis and the cumulative probability in the vertical axis. As expected, the cumulative distribution of the benchmark index lies in between the distributions of HIM and LIM at all score values. These two extreme weight allocation scenarios across lead to major variation in composite scores, and consecutively in rankings. Figure 2 is a bar plot of index scores obtained with benchmark, HIM and LIM weights. The universities are ordered based on the benchmark ranking index. Composite scores vary dramatically across the three scenarios with few exceptions.<sup>12</sup> In fact, the variation in scores is relatively lower for institutions that are high-ranked using the benchmark when compared to lower-ranked universities. In order to measure this relationship between the benchmark ranking the variation of index scores across the three different indices, we firstly calculated the coefficient of variation for each university across the three indices. We then calculated the Pearson correlation coefficient between the coefficients of variation and the benchmark index, which is -0.681 for the ARWU university rankings. It confirms the negative relation that we observe in Figure 2. Overall, we find that composite scores and/or rankings are

---

<sup>12</sup>E.g., score differences between three scenarios are less than ten for Harvard University, Stanford University, University of California, Berkeley, ESPCI ParisTech, Toulouse School of Economics, Carnegie Mellon University, City University of New York City College, Technion-Israel Institute of Technology, State University of New York Health Science Center at Brooklyn, Brandeis University, University of Texas Southwestern Medical Center, and Weizmann Institute of Science.

very sensitive to alternative weight allocations to performance criteria when analyzing the ARWU university level data.

## 4.2 ARWU Economics/Business department rankings

In this subsection, we replicate the above analysis using the ARWU EB departmental rankings. Similar to Table 7, Table 8 shows the results, which are almost equivalent to the analysis presented in the previous subsection. The only difference in the data is that instead of PCP, the agency uses the TOP indicator. In addition to measuring publications volume, it also incorporates a quality dimension by measuring the proportion of publications in top 20% Economics/Business journals. From Table 8 we can see that the HIM index is completely based on this variable. Similar to the university rankings case, the LIM index for EB departmental rankings is based only on the Alumni and Awards variables. The other variables receive zero weights.

Similar to Figures 1 and 2, we provide Figures 3 and 4 using the ARWU EB department data. Figure 3 plots the cumulative distribution functions of the three indices and, by construction, very similar to Figure 1: the density in the higher quintiles is highest for the HIM index, and lowest for LIM. As can be seen from Figure 3, almost 80% of EB departments have zero LIM scores. The same fraction have benchmark index scores between 25 and 40. Therefore, the three rankings have a different distribution of scores.

To examine the extent of rank differences among the three indices, we calculate the Spearman's rank correlation coefficient between the coefficients of variation and the benchmark index as in the previous subsection. Similar to the ARWU university rankings data, the Spearman's correlation coefficient for EB department ranking data is -0.763, suggesting that the high-ranked departments in benchmark ranking experience lower variation across the three index scores. Otherwise stated, high-ranked EB departments are less sensitive to alternative weight assignment, and lower-ranked EB departments are more sensitive.

### 4.3 THE University Rankings

The variables in Tables 7 and 8 that receive zero weights, may have a large number of zero-valued entries, which may cause concern. For instance, a majority of institutions have never had a Nobelist faculty member, and, therefore, they have a score of zero in a variable that measures the number of Nobel prizes ever attained. As a robustness check of our method, we replicate the analysis using a different data source, the THE world university rankings data. Table 9 contains the HIM and LIM weight vectors for the years 2012–2015. Again, as with ARWU data the variables are normalized, and as explained earlier, a meaningful analysis can be performed only if done separately for each year. It is worth noting that in this dataset, all the variables have non-zero entries. Yet, one variable in each case — for HIM and LIM — is assigned a zero weight (*Research* and *Citations*). Therefore, the performance of our method is not affected by the type and distribution of the variables involved.

In this subsection, we also replicate our analysis with THE dataset in order to evaluate the sensitivity of their rankings. The results are summarized in Table 9, and Figures 5 and 6. In Table 9, we can see that the HIM weights are highest for *Citations*. This result is very similar to the one obtained in the previous subsection for the ARWU university and EB department rankings, suggesting that direct research outcomes such as the publications and citations are relatively close to each other in value, and close to top ranked institutions. The LIM weights are highest for two variables: *Teaching* and *Research*. These are weighted averages of several indicators as shown in Table 5. Both are product of a subjective reputation survey and a measure of institution budget and university research income, among other variables. Similar to *Alumni* and *Awards* in ARWU data, these two variables (perceived reputation and budget) have little variation over the long-run.

Figure 5 plots the cumulative distribution functions of the benchmark, LIM and HIM indices. Differently from Figures 1 and 3, in this case the LIM index values do not have a high density on the zero-value. Instead the distribution is smooth throughout different score values. Figure 6 is a bar plot of the index scores using THE data. Similar to ARWU, there is a negative correlation between coefficient of variation and the benchmark scores suggesting that there is a lower (higher) variation in index scores for relatively higher (lower) ranked institutions: The Pearson correlation

coefficient between the coefficient of variation and the benchmark index is -0.556.

We summarize the sensitivity across rankings for all indices and for both datasets in Figures 7 and 8. To produce these bar plots we use data from only top 200 universities in both ARWU and THE, in order to set equal the number of universities. The graphs provide a bar plot of the number of universities that experience score (the former) or rank (the latter) changes in the HIM and LIM index in relation to the benchmark. It is easily observable that the majority of the universities in both ARWU and THE data experience major rank reversals. Overall, rankings are very sensitive to weight vectors used.

## 5 Conclusion and Discussion

Though highly disputed, university rankings remain impactful, popular and widely-used. An extensive literature discusses the impact of worldwide rankings in education policies, which aim for their universities to reach the worlds top-500 list. For example, Lim and Oeberg (2017) document how two quite different countries, Denmark and India, directly incorporated this target in their national policies. Lo (2014) discusses all aspects of Taiwan’s “5-year-5 billion” programme aimed to financially support their top universities in order to perform better in the world university rankings. Amsler and Bolsman (2012) describe how world university rankings have found use in policy making in other sectors of the economy. For instance, obtaining work permits in the Netherlands is relatively easier if applicants are graduates of top-ranked universities in worldwide rankings. However, if rankings are sensitive to normative choices made by ranking providers (e.g., choice of performance criteria and importance attached to these performance criteria), this defeats the purpose of the information they convey to governments, universities and individuals that rely on it to make high-stakes decisions.

The question that we ask in this paper is directly related to how sensitive these rankings are depending on the importance (weights) assigned to university performance criteria. To test the sensitivity of university rankings we adapt a popular approach in financial economics, the stochastic dominance efficiency (SDE) methodology. Using SDE we obtain two extreme case scenario rankings: HIM (highest index for most) and LIM (lowest index for most) rankings. The



SDE method uses a mixed integer programming and SDE testing to provide two very distinct ways of measuring overall university performance. HIM index has the property of stochastically dominating all other possible indices, ensuring that most universities achieve their highest possible composite scores with a set of given variables. LIM index is stochastically dominated by all other indices, ensuring that the majority of universities' index scores are at their lowest level possible.

In our empirical analysis, we compare the scores of these indices (HIM and LIM) to the benchmark index, publicly available Academic Ranking of World Universities (ARWU) and Times Higher Education (THE) indices, to test the sensitivity of both composite scores and rankings. We show that the index scores and university rankings are very sensitive to alternative weight allocations. When comparing the three types of rankings (benchmark, HIM, LIM), only seven universities' index scores vary by less than 10 points (note that composite scores range between 0 and 100). The rest vary by more than 10 points. Similar to the composite score variation, rankings of THE and ARWU are also sensitive to alternative weight allocations across performance criteria. We also analyze the relative sensitivity of index scores and ranks for different groups of universities, and find that scores and ranks of top-ranked universities are less sensitive to alternative-weight allocations when compared to the lower ranked universities. Therefore, rankings of top-ranked universities are relatively more robust to weight allocations than low-ranked universities.

The sensitivity of index scores and ranks are by-products of the performance criteria used by ranking providers. For example, low- and middle-ranked universities in ARWU and THE would move to higher positions if one were to use publications indicators in ARWU and research influence (i.e., citations) indicator in THE (heavily weighted in HIM scenario) since the majority of universities have relatively high scores in these indicators. The reverse is true if some other performance criteria are heavily weighted. Those performance criteria include the number of Nobel prizes won by alumni and staff members in ARWU rankings, and teaching and research reputation, and institutional income variables in the THE university rankings (heavily weighted in LIM scenario). These variables build inertia into rankings that is of benefit to older and consistently high-ranked institutions. They are characterized by a relatively higher variance, and this may be a reason why composite scores and rankings are very sensitive to alternative-weight

allocations. Therefore, classification of performance criteria into different groups, which would lead to less volatility in composite scores and rankings, may be useful in building more robust university rankings.

Finally, it is worth noting that this paper does not aim to provide any theoretical discussion on what should constitute a better measurement of university quality. This would require a systematic analysis that would involve stakeholders in the selection of performance criteria, their measurement, and aggregation methodology. What our paper does is to examine the sensitivity of university rankings based on choices made (on performance criteria and their weights) in the construction of two world university rankings, ARWU and THE. Our methodology can be used to assess the sensitivity of other institutional rankings in different industries.

# References

- Agliardi, E., Agliardi, R., Pinar, M., Stengos, T., Topaloglou (2012). A new country risk index for emerging markets: A stochastic dominance approach. *Journal of Empirical Finance*, 19(5), 741-761.
- Agliardi, E., Pinar, M., Stengos, T. (2014). A sovereign risk index for the Eurozone based on stochastic dominance. *Finance Research Letters*, 11(4), 375-384.
- Agliardi, E., Pinar, M., Stengos, T. (2015). An environmental degradation index based on stochastic dominance. *Empirical Economics*, 48(1), 439-459.
- Agliardi, E., Pinar, M., Stengos, T. (2017). Air and water pollution over time and industries with stochastic dominance. *Stochastic Environmental Research and Risk Assessment*, 131(6), 1389-1408.
- Amsler, S. S., and Bolsman, C., (2012). University ranking as social exclusion. *British Journal of Sociology of Education*, 33:2, 283–301.
- Atkinson, A.B. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, 2, 244-263.
- Atkinson, A.B. (1987). On the Measurement of Poverty. *Econometrica*, 55(41), 749-764.
- Barrett, G., Donald, S. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71, 71-104.
- Bazen, S., Moyes, P. (2012). Elitism and stochastic dominance. *Social Choice and Welfare*, 39 (1), 207-251.
- Billaut, J.-C., Bouyssou, D., Vincke, P. (2010). Should you believe in the Shanghai ranking?: An MCDM view. *Scientometrics*, 84(1), 237-263.
- Bowman, N.A., Michael N. Bastedo, M.N. (2011). Anchoring effects in world university rankings: exploring biases in reputation scores. *Higher Education*, 61(4), 431-444.
- Broecke, S. (2015). University rankings: do they matter in the UK? *Education Economics*, 23(2), 137-161.
- Carayol, N. and Lahatte, A. (2014). "Dominance relations and ranking when quantity and quality both matter: Applications to US universities and econ. departments worldwide," *Cahiers du*

GREThA 2014-14, GGroupe de Recherche en Economie Théorique et Appliquée. URL: <https://ideas.repec.org/14.html>

Cyrenne, P., Grant, H. (2009). University decision making and prestige: An empirical study. *Economics of Education Review*, 28, 237-248.

Davidson, R., Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68, 1435-1464.

Delgado, M.A., Escanciano, J.C. (2013). Conditional stochastic dominance testing. *Journal of Business and Economic Statistics*, 31(1), 16-28.

Docampo, D., Egret, D., Cram, L. (2015). The effect of university mergers on the Shanghai ranking. *Scientometrics*, 104, 175-191.

Duclos, J.-Y., Sahn, D.E., Younger, S.D. (2006). Robust multidimensional poverty comparisons. *Economic Journal*, 116(514), 943-968.

Foster, J.E., Shorrocks, A. (1988). Poverty Orderings and Welfare Dominance. *Social Choice and Welfare*, 5(2-3), 179-198.

Gibbons, S., Neumayer, E., Perkins, R. (2015). Student satisfaction, league tables and university applications: Evidence from Britain. *Economics of Education Review*, 48, 148-164.

Gonzalo, J., Olmo, J. (2014). Conditional stochastic dominance tests in dynamic settings. *International Economic Review*, 55(3), 819-838.

Griffith, A., Rask, K. (2007). The influence of the US News and World Report collegiate rankings on the matriculation decision of high-ability students: 1995-2004. *Economics of Education Review*, 26, 244-255.

Hazelkorn, E. (2011). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Houndsmills, UK: Palgrave MacMillan.

Hazelkorn, E. (2014). Reflections on a Decade of Global Rankings: what we've learned and outstanding issues. *European Journal of Education*, 49(1), 12-28.

Horstschräer, J. (2012). University rankings in action? The importance of rankings and an excellence competition for university choice of high-ability students. *Economics of Education Review*, 31, 1162-1176.

- Kakwani N. (1984). Welfare ranking of income distributions. *Advances in Econometrics*, 3, 191-213.
- Kehm, B.M. (2014). Global University Rankings - Impacts and Unintended Side Effects. *European Journal of Education*, 49(1), 102-112.
- Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10), 1390-1406.
- Lim, M. A., and Oeberg, J. W. (2017). Active instruments: on the use of university rankings in developing national systems of higher education. *Policy Reviews in Higher Education*, 1(1), 91-108.
- Linton, O., Maasoumi E., Whang, Y.-J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, 72(3), 735-765.
- Linton, O., Post, T., Whang, Y.-J. (2014). Testing for the stochastic dominance efficiency of a given portfolio. *Econometrics Journal*, 17, 59-74.
- Lo, W.Y.W. (2014). *University Rankings*. Springer Science+Business Media Singapore. DOI: 10.1007/978-981-4560-35-1\_4.
- Milla, J. (2017). The Context-Bound University Selectivity Premium. IZA-Institute of Labor Economics Discussion Paper No. 11025. URL: <http://ftp.iza.org/dp11025.pdf>
- Pinar, M. (2015). Measuring world governance: revisiting the institutions hypothesis. *Empirical Economics*, 48(2), 747-778.
- Pinar, M., Stengos, T., Topaloglou, N. (2013). Measuring human development: a stochastic dominance approach. *Journal of Economic Growth*, 18(1), 69-108.
- Pinar, M., Stengos, T., Yazgan, M.E. (2015). Measuring human development in the MENA region. *Emerging Markets Finance and Trade*, 51(6), 1179-1192.
- Pinar, M., Stengos, T., Topaloglou, N. (2017). Testing for the implicit weights of the dimensions of the Human Development Index using stochastic dominance. *Economics Letters*, 161, 38-42.
- Post, T. (2003). Empirical tests for stochastic dominance efficiency. *Journal of Finance*, 58, 1905-1931.
- Ravallion, M. (1994). *Poverty Comparisons, Fundamentals of Pure and Applied Economics*.

Harwood Academic Publishers, Switzerland.

Rio E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendants. Mathématiques et Applications, 31. Springer, Berlin.

Saisana, M., d'Hombres, B., Saltelli, A. (2011). Rickety numbers: Volatility of university rankings and policy implications. Research Policy, 40, 165-177.

Scaillet, O., Topaloglou, N. (2010). Testing for stochastic dominance efficiency. Journal of Business and Economic Statistics, 28(1), 169-180.

Shorrocks A.F. (1983). Ranking income distributions. Economica, 50(197), 3-17.

Torres-Salinas, D., Moreno-Torres, J.G., Emilio Delgado-Lopez-Cozar, E., Herrera, F. (2011). A methodology for Institution-Field ranking based on a bidimensional analysis: the  $IFQ^2A$  index. Scientometrics, 88, 771-786.

Yalonetzky, G. (2014). Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables. Social Choice and Welfare, 43(4), 773-807.

## Appendix

Table 1. Variables of the ARWU university ranking data and pre-determined weight allocations		
Performance criteria	Criteria code	Weight
The total number of alumni of an institution winning Nobel prizes and fields medals	Alumni	10%
The total number of the staff of an institution winning Nobel Prizes in Physics, Chemistry, Medicine and Economics and Fields Medal in Mathematics	Award	20%
The number of Highly Cited Researchers selected by Thomson Reuters	HiCi	20%
The total number of papers published in Nature and Science	N&S	20%
Total number of papers indexed in Science Citation Index-Expanded and Social Science Citation Index	PUB	20%
The weighted scores of the other five indicators (Alumni, Award, HiCi, N&S, and PUB) divided by the number of full-time equivalent academic staff.	PCP	10%

Table 2. Variables of the ARWU economics/business department data and pre-determined weight allocation		
Performance criteria	Criteria code	Weight
Total number of alumni of an institution winning Nobel Prizes in Economics since 1961	Alumni	10%
Total number of staff of an institution winning Nobel Prizes in Economics since 1971	Award	15%
Total number of staff of an institution that is listed in the highly cited researchers in Economics/Business Category	HiCi	25%
Total number of papers Indexed in Social Science Citation Index in Economics/Business fields	PUB	25%
Percentage of papers published in top 20% journals of Economics/Business fields	TOP	25%

Table 3. Descriptive statistics of the ARWU worldwide university rankings data						
Year\Indicator	<i>Alumni</i>	<i>Award</i>	<i>HiCi</i>	<i>N&amp;S</i>	<i>PUB</i>	<i>PCP</i>
<b>2015</b>						
Mean	7.96	7.43	15.25	15.29	38.85	21.79
Median	0.00	0.00	12.30	12.10	36.75	19.90
Standard Dev.	12.69	15.39	13.20	11.87	12.47	8.77
Skewness	2.56	3.03	1.81	2.30	0.86	3.44
<b>2014</b>						
Mean	8.04	7.22	15.22	15.85	38.95	21.42
Median	0.00	0.00	12.20	12.30	36.90	19.70
Standard Dev.	12.78	15.15	13.18	12.27	12.48	8.72
Skewness	2.57	3.06	1.83	2.26	0.90	3.45
<b>2013</b>						
Mean	8.07	7.17	15.97	15.45	38.03	20.43
Median	0.00	0.00	12.50	12.20	35.70	18.70
Standard Dev.	12.79	15.08	14.27	12.12	12.24	8.56
Skewness	2.57	3.07	1.74	2.26	0.92	3.65
<b>2012</b>						
Mean	8.06	7.20	15.97	15.66	38.01	19.84
Median	0.00	0.00	12.50	12.50	36.00	18.10
Standard Dev.	12.85	15.25	14.28	12.12	12.30	8.41
Skewness	2.55	3.08	1.75	2.27	0.95	3.73
<b>2011</b>						
Mean	8.63	7.25	15.90	15.66	37.87	19.90
Median	0.00	0.00	12.50	12.70	35.50	18.10
Standard Dev.	13.69	15.04	14.28	12.10	12.46	8.34
Skewness	2.44	3.04	1.74	2.29	0.95	3.65
<b>2010</b>						
Mean	8.55	7.01	15.64	15.20	38.12	20.24
Median	0.00	0.00	12.50	12.30	35.85	18.40
Standard Dev.	13.56	14.78	14.12	11.99	12.79	8.37
Skewness	2.47	3.09	1.78	2.34	0.95	3.60
<b>2009</b>						
Mean	8.59	6.91	15.64	14.93	37.32	21.31
Median	0.00	0.00	12.60	11.90	35.20	19.00
Standard Dev.	13.68	14.73	14.01	11.82	12.75	9.39
Skewness	2.46	3.11	1.81	2.40	0.97	2.91
<i>Note:</i> Numbers of institutions covered in 2015, 2014, 2013, 2012, 2011, 2010 and 2009 are 498, 497, 498, 497, 497, 498, and 499 respectively.						



Table 4. Descriptive statistics of the ARWU economics/business department data					
Year\Indicator	<i>Alumni</i>	<i>Award</i>	<i>HiCi</i>	<i>PUB</i>	<i>TOP</i>
<b>2015</b>					
Mean	5.68	5.75	10.90	49.69	72.84
Median	0.00	0.00	9.10	47.20	72.45
Standard Dev.	16.26	15.45	15.70	13.00	12.29
Skewness	3.11	3.15	2.32	0.82	-0.04
<b>2014</b>					
Mean	5.65	5.50	10.58	49.90	72.18
Median	0.00	0.00	9.10	47.65	71.65
Standard Dev.	17.70	15.14	15.69	13.13	12.80
Skewness	3.32	3.28	2.38	0.87	-0.03
<b>2013</b>					
Mean	5.57	5.90	13.20	48.35	73.86
Median	0.00	0.00	0.00	46.05	74.30
Standard Dev.	17.01	16.12	18.53	13.46	11.57
Skewness	3.23	3.17	1.79	0.85	-0.31
<b>2012</b>					
Mean	5.19	5.65	13.20	46.60	71.12
Median	0.00	0.00	0.00	43.85	70.60
Standard Dev.	16.38	15.76	18.53	13.05	12.25
Skewness	3.42	3.28	1.79	0.99	-0.10
<b>2011</b>					
Mean	10.14	10.39	23.09	52.30	77.25
Median	0.00	0.00	17.80	49.00	78.70
Standard Dev.	23.50	20.11	20.46	13.64	10.38
Skewness	2.35	2.16	1.22	0.75	-0.47
<b>2010</b>					
Mean	11.13	9.89	23.19	51.13	73.95
Median	0.00	0.00	17.80	47.60	75.30
Standard Dev.	23.77	19.94	19.98	12.80	9.75
Skewness	2.13	2.24	1.24	0.87	-0.31
<b>2009</b>					
Mean	11.22	9.71	23.18	50.76	76.40
Median	0.00	0.00	18.00	48.00	77.50
Standard Dev.	23.28	19.64	19.60	13.65	10.78
Skewness	2.07	2.27	1.27	0.83	-0.30
Note: Numbers of institutions covered between the periods of 2011-2015 and 2009-2011 are 200 and 100 respectively.					

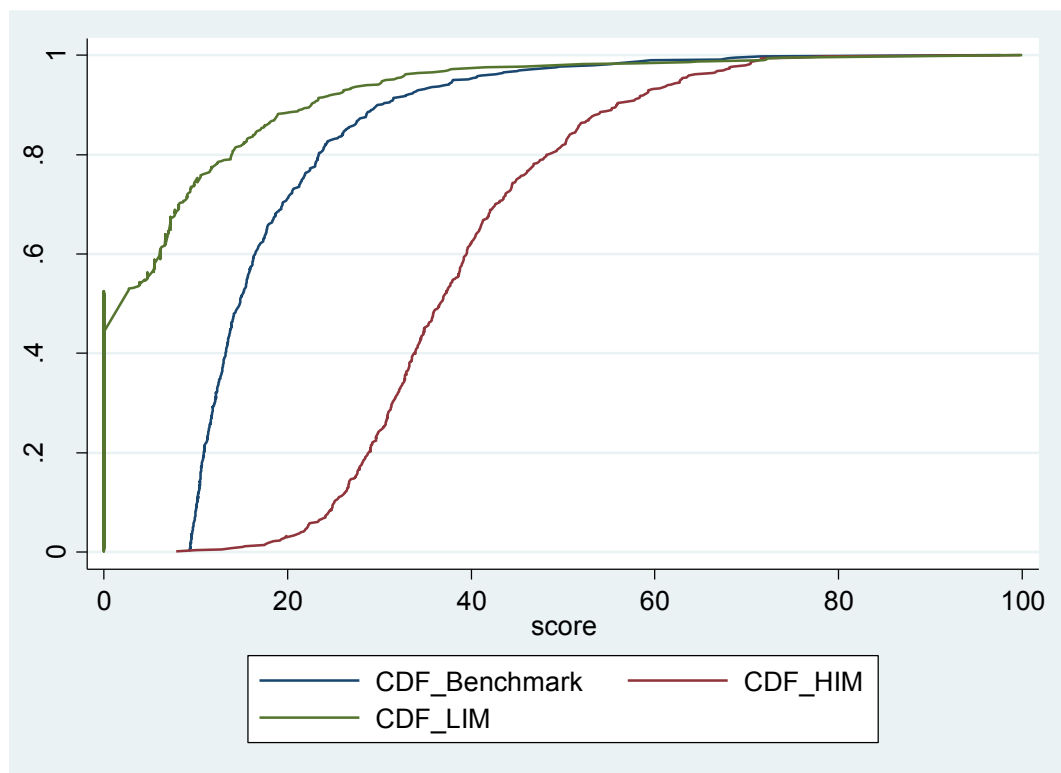
Table 5. Variables of the THE world university ranking data and pre-determined weight allocations			
Performance criteria	Indicators and their measurement	Criteria weight	Indicator weight
Teaching	Perceived prestige of institutions in teaching based on Thomson Reuters' reputation survey of worldwide poll of experienced scholars	30%	15%
	Staff-to-student ratio		4.5%
	Ratio of doctoral to bachelor's degrees awarded by each institution.		2.25%
	The number of doctorates awarded by an institution, scaled against its size as measured by the number of academic staff it employs		6%
	Measure of institutional income scaled against academic staff numbers		2.25%
Research	A university's reputation for research excellence among its peers, based on the 10,000-plus responses to annual academic reputation survey	30%	18%
	University research income, scaled against staff numbers and normalised for purchasing-power parity		6%
	The number of papers published in the academic journals indexed by Thomson Reuters per academic, scaled for a university's total size and also normalised for subject		6%
Citations	This dimensions measures the number of times a university's published work is cited by scholars globally. The data are drawn from the 12,000 academic journals indexed by Thomson Reuters' Web of Science database and include all indexed journals published over five years (depending on the evaluation of the year), and counts the citations to these papers made in the six years. The data are fully normalised to reflect variations in citation volume between different subject areas	30%	30%
Industry	Research income an institution earns from industry, scaled against the number of academic staff it employs	2.5%	2.5%
International outlook	The ratio of international to domestic students	7.5%	2.5%
	The ratio of international to domestic staff		2.5%
	The proportion of a university's total research journal publications that have at least one international co-author and reward higher volumes. This indicator is normalised to account for a university's subject mix and uses the same five-year window as the "Citations" dimension.		2.5%

Table 6. Descriptive statistics for the Times Higher Education data					
Criteria	Teaching	Research	Citations	Industry	International Outlook
<i>Descriptive statistics for 2015</i>					
Mean	38.06	37.12	68.23	51.19	56.68
Median	34.10	32.40	68.20	43.40	55.60
Standard deviation	15.97	19.63	16.65	21.08	19.56
Skewness	1.33	1.18	-0.03	1.07	0.11
<i>Descriptive statistics for 2014</i>					
Mean	36.78	35.58	66.22	50.82	55.60
Median	32.70	30.70	65.80	42.80	55.70
Standard deviation	16.62	19.36	17.32	20.81	20.01
Skewness	1.35	1.31	-0.02	1.11	0.09
<i>Descriptive statistics for 2013</i>					
Mean	41.22	40.79	64.50	50.23	53.44
Median	37.30	35.75	63.55	42.30	52.80
Standard deviation	18.18	21.49	18.51	20.80	20.36
Skewness	0.94	0.89	0.06	1.17	0.12
<i>Descriptive statistics for 2012</i>					
Mean	36.98	35.04	56.20	46.95	52.07
Median	32.50	30.00	54.40	38.60	50.40
Standard deviation	18.19	20.91	24.03	21.90	23.05
Skewness	1.15	1.14	0.13	1.23	0.14
<i>Notes:</i> There are 381, 367, 364, and 369 institutions that have information for all dimensions used for the Times Higher Education world university rankings in 2015, 2014, 2013 and 2012, respectively.					

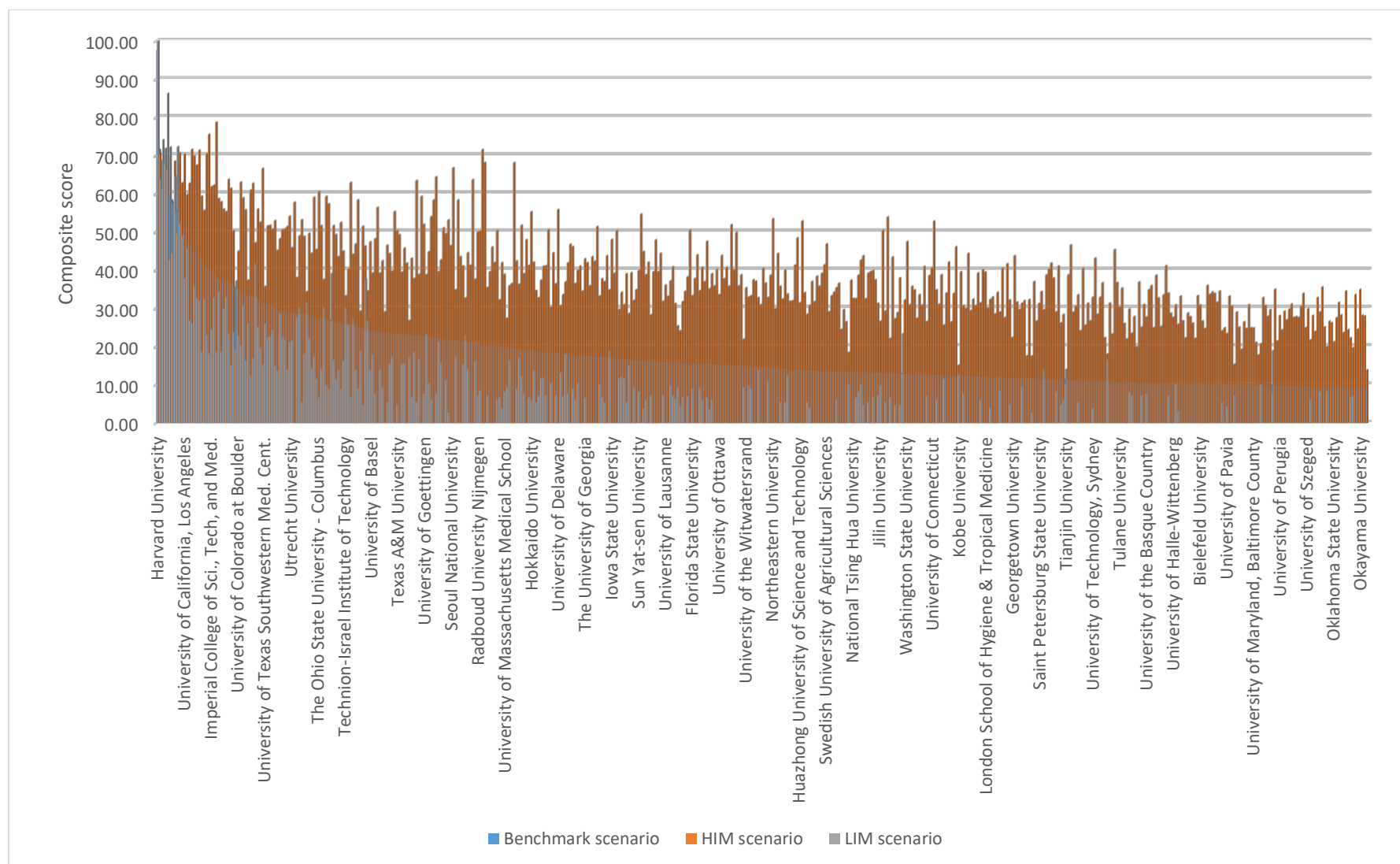
Table 7. HIM and LIM weight allocations using the ARWU university ranking data						
Panel A. Weight allocation across the indicators for the HIM scenario						
Year\Indicator	<i>Alumni</i>	<i>Award</i>	<i>HiCi</i>	<i>N&amp;S</i>	<i>PUB</i>	<i>PCP</i>
Period	Average weight allocations					
<b>2015</b>	0.000	0.000	0.000	0.000	0.992	0.008
<b>2014</b>	0.000	0.000	0.000	0.000	0.985	0.015
<b>2013</b>	0.000	0.000	0.000	0.000	0.982	0.018
<b>2012</b>	0.000	0.000	0.000	0.000	0.998	0.002
<b>2011</b>	0.000	0.000	0.000	0.000	0.995	0.005
<b>2010</b>	0.000	0.000	0.000	0.000	0.992	0.008
<b>2009</b>	0.000	0.000	0.000	0.000	0.997	0.003
Panel B. Weight allocation across the indicators for the LIM scenario						
Year\Indicator	<i>Alumni</i>	<i>Award</i>	<i>HiCi</i>	<i>N&amp;S</i>	<i>PUB</i>	<i>PCP</i>
Period	Average weight allocations					
<b>2015</b>	0.530	0.470	0.000	0.000	0.000	0.000
<b>2014</b>	0.462	0.538	0.000	0.000	0.000	0.000
<b>2013</b>	0.442	0.558	0.000	0.000	0.000	0.000
<b>2012</b>	0.452	0.548	0.000	0.000	0.000	0.000
<b>2011</b>	0.362	0.638	0.000	0.000	0.000	0.000
<b>2010</b>	0.383	0.617	0.000	0.000	0.000	0.000
<b>2009</b>	0.326	0.674	0.000	0.000	0.000	0.000
<i>Note:</i> Numbers of institutions covered in 2015, 2014, 2013, 2012, 2011, 2010 and 2009 are 498, 497, 498, 497, 497, 498, and 499 respectively.						

Table 8. HIM and LIM weight allocations using the ARWU economics/business departments data					
Panel A. Weight allocation across the indicators for the HIM scenario					
Period	<i>Alumni</i>	<i>Award</i>	<i>HiCi</i>	<i>PUB</i>	<i>TOP</i>
Period	Average weight allocations				
<b>2015</b>	0.000	0.000	0.000	0.000	1.000
<b>2014</b>	0.000	0.000	0.000	0.000	1.000
<b>2013</b>	0.000	0.000	0.000	0.000	1.000
<b>2012</b>	0.000	0.000	0.000	0.000	1.000
<b>2011</b>	0.000	0.000	0.000	0.000	1.000
<b>2010</b>	0.000	0.000	0.000	0.000	1.000
<b>2009</b>	0.000	0.000	0.000	0.000	1.000
Panel B. Weight allocation across the indicators for the LIM scenario					
Period	<i>Alumni</i>	<i>Award</i>	<i>HiCi</i>	<i>PUB</i>	<i>TOP</i>
Period	Average weight allocations				
<b>2015</b>	0.418	0.582	0.000	0.000	0.000
<b>2014</b>	0.435	0.565	0.000	0.000	0.000
<b>2013</b>	0.471	0.529	0.000	0.000	0.000
<b>2012</b>	0.447	0.553	0.000	0.000	0.000
<b>2011</b>	0.334	0.636	0.000	0.000	0.000
<b>2010</b>	0.472	0.528	0.000	0.000	0.000
<b>2009</b>	0.408	0.592	0.000	0.000	0.000
<i>Note:</i> Numbers of institutions covered between the periods of 2011-2015 and 2009-2011 are 200 and 100, respectively.					

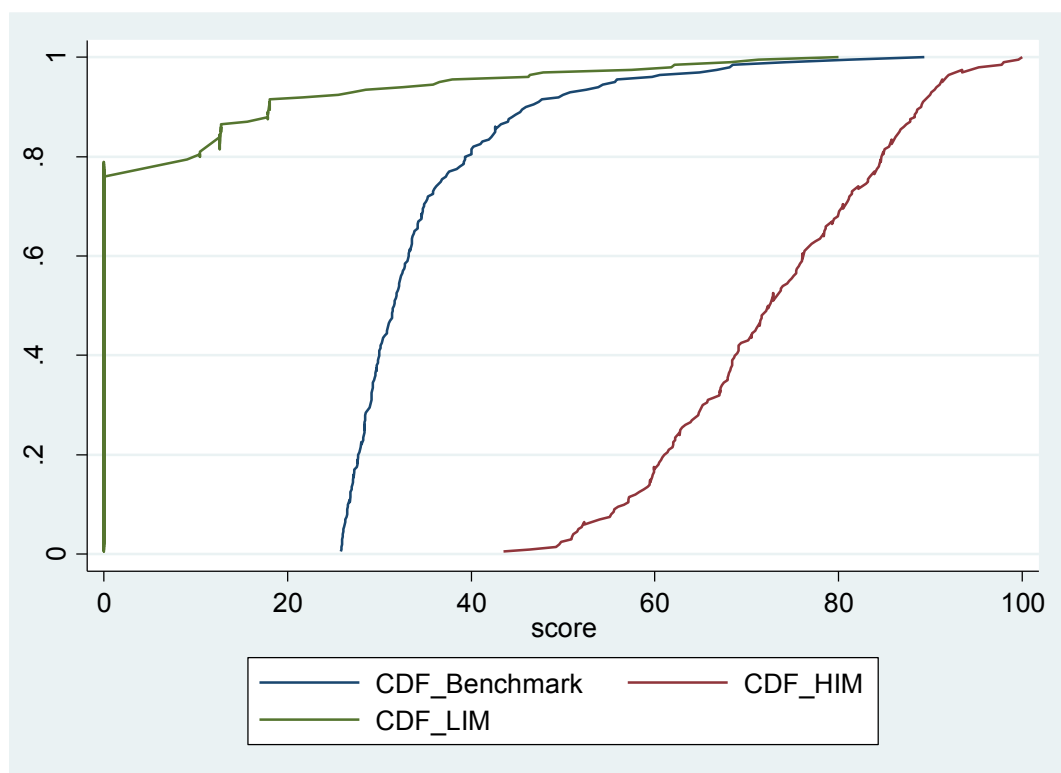
Table 9. HIM and LIM weight allocations using the Times Higher Education data					
Panel A. Weight allocation across the criteria for the HIM scenario					
	<i>Teaching</i>	<i>Research</i>	<i>Citations</i>	<i>Industry</i>	<i>International outlook</i>
Period	Average weight allocations				
<b>2015</b>	0.060	0.000	0.738	0.136	0.066
<b>2014</b>	0.047	0.000	0.691	0.179	0.083
<b>2013</b>	0.052	0.000	0.690	0.160	0.098
<b>2012</b>	0.081	0.000	0.483	0.195	0.241
Panel B. Weight allocation across the criteria for the LIM scenario					
	<i>Teaching</i>	<i>Research</i>	<i>Citations</i>	<i>Industry</i>	<i>International outlook</i>
Period	Average weight allocations				
<b>2015</b>	0.511	0.425	0.000	0.029	0.035
<b>2014</b>	0.451	0.488	0.000	0.032	0.029
<b>2013</b>	0.476	0.397	0.000	0.047	0.080
<b>2012</b>	0.302	0.635	0.000	0.022	0.041
<i>Notes:</i> Numbers of institutions covered in 2015, 2014, 2013, and 2012 are 381, 367, 364, and 369, respectively.					



**Figure 1.** Cumulative distributions of the 2015 ARWU university benchmark, HIM and LIM index scores

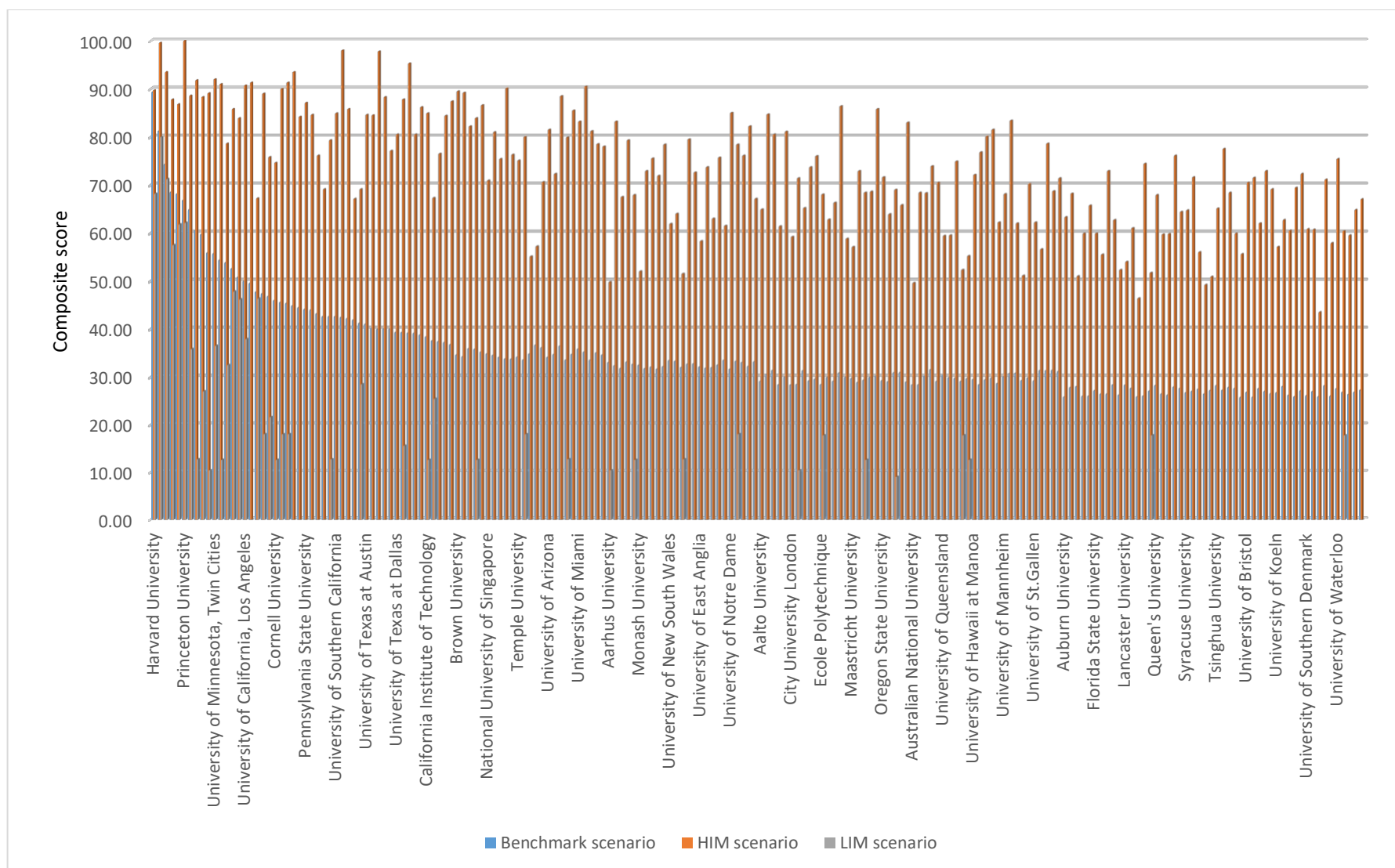


**Figure 2.** Plot of the 2015 ARWU university benchmark, HIM and LIM index scores

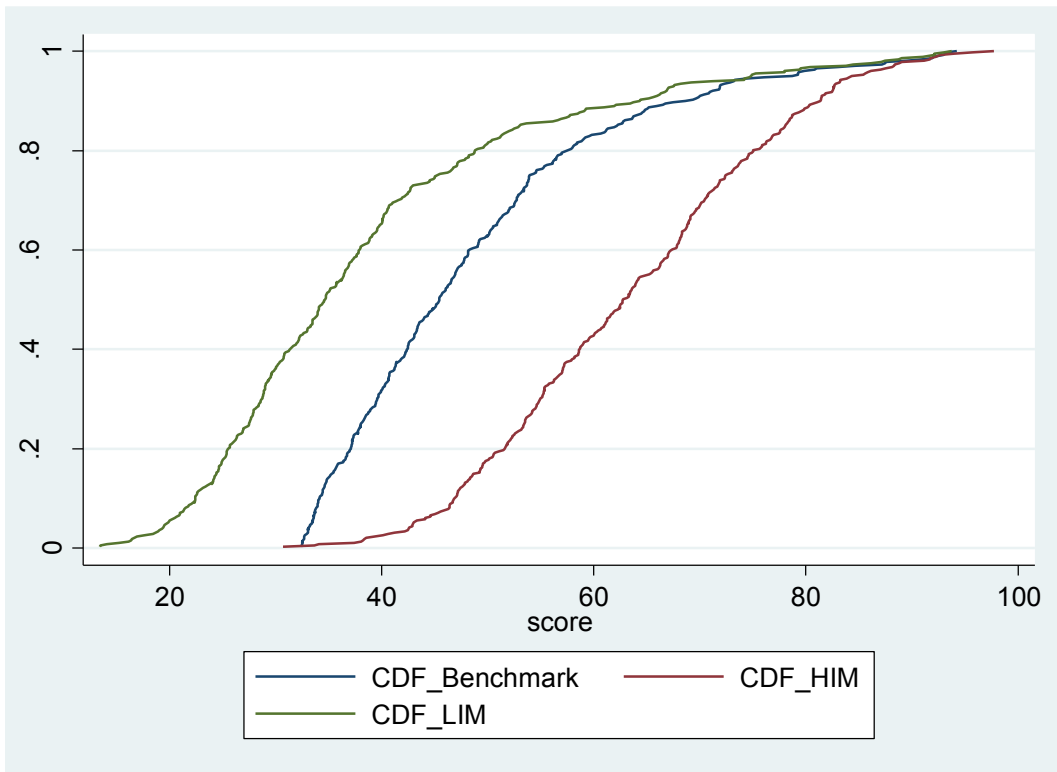


**Figure 3.** Cumulative distributions of the 2015 ARWU economics/business department benchmark, HIM and LIM index scores

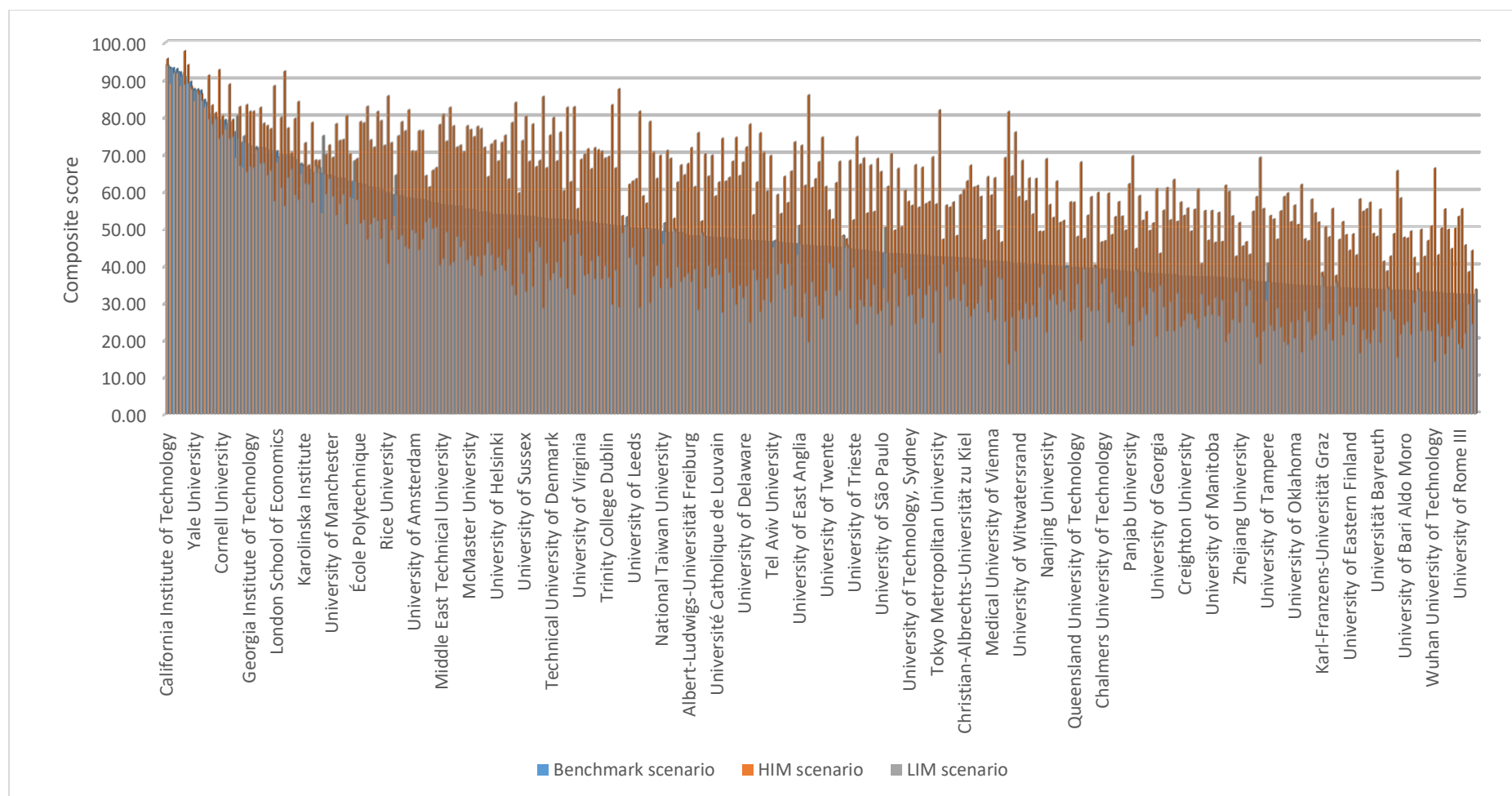




**Figure 4.** Plot of the 2015 ARWU business/economics department benchmark, HIM and LIM index scores



**Figure 5.** Cumulative distributions of the 2015 THE university benchmark, HIM and LIM index scores



**Figure 6.** Plot of the 2015 THE university benchmark, HIM and LIM index scores

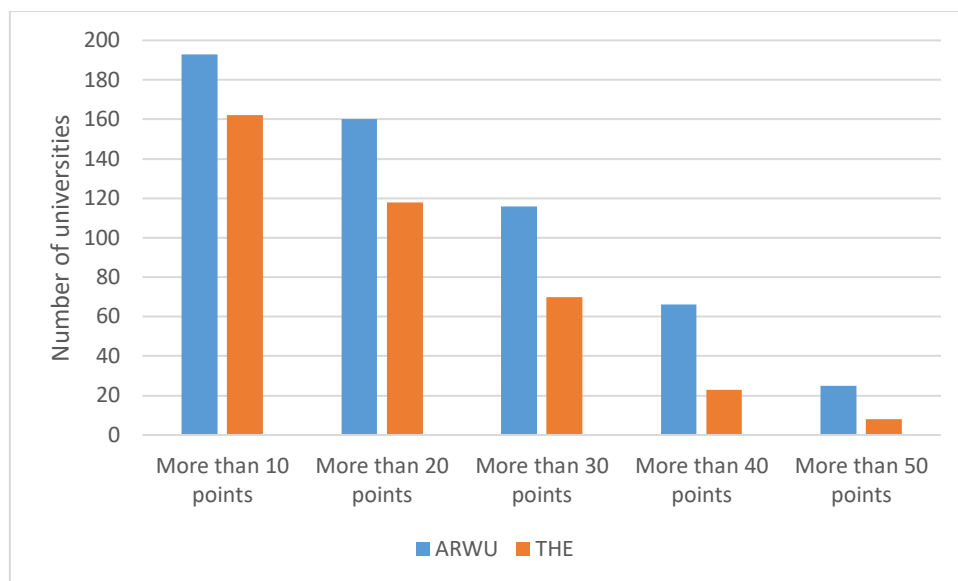


Figure 7. Distribution of composite score changes between benchmark versus HIM and LIM indices

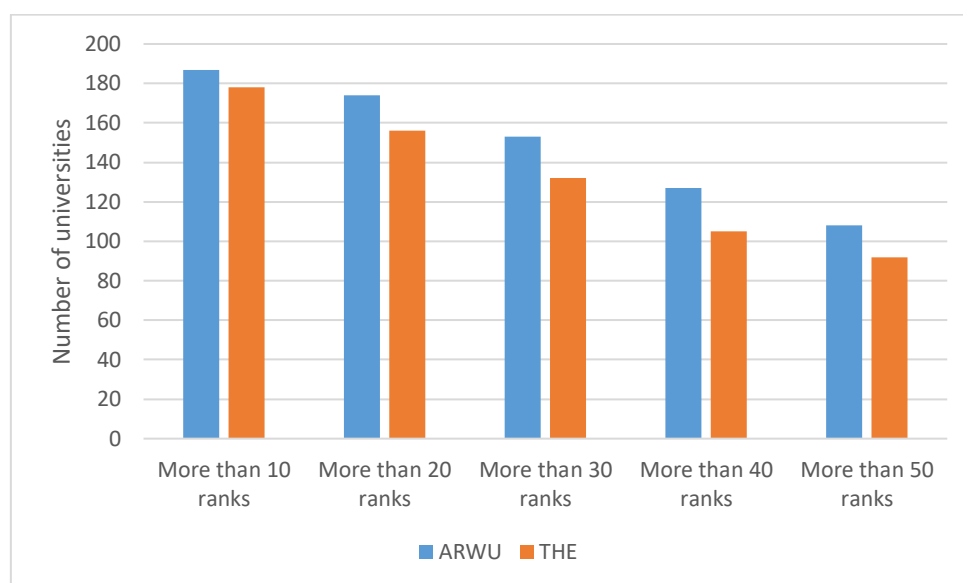


Figure 8. Distribution of rank changes between benchmark versus HIM and LIM indices